Document: NBD-WG N0002

Minutes: NIST Big Data Working Group (NBD-WG) Teleconference
June 26, 2013, 13:00 - 15:00 EDT

Chris Greer (Associate Director for. Program Implementation in NIST's ITL) opened the meeting by welcoming participants and providing a brief introduction emphasizing the goal of obtaining the "best consensus available now" on Big Data topics.

Bob Marcus and Chaitan Baru (NBD-WG Co-Chairs) gave brief opening remarks emphasizing the tight schedule and the need for urgency in developing deliverables.

Wo Chang (NIST Leader for the NBD-WG) presented an overview of the Agenda, Charter, and Standing Rules for the Working Group as below:

Standing Rules
- All information exchanged within the WG will be non-proprietary.
- All information exchanged within the WG will contain non-PII materials.
- WG members should assume that all materials exchanged will be made public.
- Documents will be publicly accessible on the NIST Big Data Portal.

Bob suggested started subgroup activities (e.g. mailing list) as soon as possible. He also suggested that the full NBD-WG meet less frequently than once a week to enable more time for subgroup activities. Wo agreed with less frequent full WG meetings. He announced that the subgroup collaboration tools would be available shortly.

An attendee suggested there was a need to define scope of the subgroups. Wo will send some initial subgroup descriptions. Subgroups can develop their detailed charters.

There were some questions concerning mailing lists vs. collaboration Web site and subgroups vs reflector that were addressed by Wo.

An attendee asked about the deliverable from the Technology Roadmap subgroup. Wo explained that this deliverable would be similar to the Cloud Technology Roadmap document.

There was a suggestion from the audience that the subgroups use common working processes and tools to simplify participation in multiple subgroups. Wo agreed that this was a good idea.

An attendee asked where use cases will be discussed Wo replied that the Requirements subgroup would extract requirements from use cases.

An attendee asked where ontologies would be discussed. Wo replied that this topic should be part of the Taxonomy subgroup

An attendee asked where do deliverables (e.g. Big Data Definition) go. Wo replied that deliverables can feed into other subgroups asynchronously. Discussions among subgroups can help with coordination. The Technology Roadmap subgroup can resolve overlaps.

Chaitan said that a collaborative Web space was needed. He also recommended that NIST Cloud documented should be reviewed to provide guidance to Big Data subgroups. Wo will post Cloud documents.

Wo said that subgroup documents should be posted by the author and then discussed on reflector. Authors should post revisions. Chaitan emphasized that the same process should be followed by all subgroups.

An audience member asked how documents would be posted. Wo said that anyone could submit documents and then demonstrated the upload interface. He said that subgroup co-chairs could determine schedule for meetings.

Chaitan suggested a training session for subgroup co-chairs. Wo will post subgroup weekly meeting times and set up video conferencing rooms for each subgroup.

An attendee pointed out the dependency among subgroup deliverables. He suggested that the scope should be clear and information shared across groups.

There was a discussion about the number of subgroups. There was some audience concern that there might be too many subgroups. Wo believes that multiple subgroups can accelerate progress but we have to plan how to bring the subgroup deliverables together. He also stated that government agencies are aware of NIST's Big Data activity and that the Working Group should not "reinvent the wheel".

An audience member asked about the customers for the deliverables. Wo said it was similar to the NIST Cloud deliverables. It could be the end-users or solution providers.

An audience member noted that there diverse models for different users and multiple perspectives by vendors. Wo stated that we use existing work (e.g. Reference Architectures) but remain technology neutral.

An audience member express concern over the timeframe for deliverables. She asked about the maturation plan and noted that the Cloud Working Groups had encountered unexpected delays. Wo fully agreed that this was a key challenge that had to be addressed.

Wo provide a recap. He reminded all participants to upgrade their profiles to indicate subgroups of interest. He will provide high level descriptions of subgroups activities and expected deliverables. People who want to be considered for Co-chairs should send him e-mail.

Appendix: Web Chat Session Comments

(10:09 AM) Karen Guertler: An observation: The list of deliverables in section IV b appears to be different from the list of current work groups & 'reflectors'.

(10:15 AM) Karen Guertler: I'm quite interested in how this NIST initiative relates to other initiatives around big data - especially wrt NIH.

(10:15 AM) Tom Plunkett: Will all of the sub-groups have all of their own scheduled call times, or will you subdivide this call time for the different sub groups

(10:17 AM) Karen Guertler: Will there also be oppties to collaborate using technologies beyond e-mail? for example, via wiki's or SharePoint?

(10:19 AM) Geoffrey Fox (Indiana University): The registration site had a use case wg but I didn't see that on your list

(10:20 AM) Dr. Ravi Sharma: Does there have to be one definition for all purposes? or there are possibilities of differentiation by domain, context etc.?

(10:20 AM) Dave Raddatz (SGI): what is a reflector?

(10:21 AM) Dr. Ravi Sharma: Why are we not having Ontology of big Data as i was co-chair at NIST Cloud Workshop which had this topic, will it be under Taxonomy but ontology overarches taxonomy?

(10:24 AM) Karen Guertler: If so, excellent question, as the taxonomies would vary by industry, correct?

(10:24 AM) Anil Srivastava (OHSL/ICTBioMed): We have an international consortium of supercomputers working on next general cyber capability with emphasis on big data. Is it possible for them to participate in these discussions?

(10:25 AM) Carl Buffington (USDA NRCS): Sounds like we need a charter for each subgroup?

(10:25 AM) Dr. Ravi Sharma: suggest name as Ontology and taxonomy

(10:25 AM) Karen Guertler: Carl, I agree.

(10:25 AM) Scott Brim (Internet2): Excuse me, but WHERE will you send the list of subgroups? Which list?

(10:26 AM) _James Kobielus_IBM: please set up a wiki so we can ask the questions and get detailed answers and discuss things coherently there...this audio is incredibly noisy and distracting

(10:26 AM) Janis Bech (IMS): Need to use a webinar like Web Ex with call in all muted unless you raise your hand and are switched to speaking mode. The background noise is difficult.

(10:27 AM) Karen Guertler: I agree, and believe a collaborative space would be preferable to an e-mail list.

(10:27 AM) Dr. Ravi Sharma: will you be able to review comments on chat? Dr. Wo.

(10:28 AM) Tom Plunkett: +1 to wiki as better than email

(10:29 AM) Orit Levin (Microsoft): The document upload link seems to be not activated. When will it be activated or is there an alternative document submission tool?

(10:29 AM) Karen Guertler: @Tom thanks! wiki or any other collaborative space.

(10:30 AM) Tom Plunkett: look at example roadmap and ref arch as used for cloud

(10:30 AM) Karen Guertler: @IBM - completely agree - need to define the audience & objectives.
(10:30 AM) Tom Plunkett: See nist.gov/itl/csd/cloud-091311.cfm for cloud version of both
(10:32 AM) Linda Pelekoudas: So what you are saying is that this "Roadmap" is a sumarization of common terms and best practices. Is this correct?
(10:33 AM) Karen Guertler: I'll suggest that a first objective should be to clarify objectives & goals of this initiative.
(10:33 AM) Bob Marcus (ET-Strategies): I think that we need collaborative Web sites for subgroups as soon as possible.
(10:34 AM) Keith W. Hare (JCC Consulting, Inc.): Take a look at http://bigdatawg.nist.gov/NBD-WG_workplan_v1.pdf for a better idea of the the definition of the objectives.
(10:35 AM) William Miller (MaCT USA): Should we not use the Cloud Infrstructure including the Service Broker
(10:35 AM) William Miller (MaCT USA): The Service Broker can make the decision to recover data from a repository
(10:35 AM) Pw Carey, (Compliance Partners, LLC): We're interested in using Big Data for conducting Digital Forensics within the Cloud Eco-System.....?
(10:35 AM) William Miller (MaCT USA): we should review the cloud standard document
(10:36 AM) Bob Marcus (ET-Strategies): We should also have a collaborative Web site for the full Working Group
(10:36 AM) Pw Carey, (Compliance Partners, LLC): In the future, can we use SOAP HUB....?
(10:37 AM) Dr. Ravi Sharma: It should be other way around Onto and Taxo
(10:38 AM) Dr. Ravi Sharma: I have not understood how hierachies are under Taxonomy
(10:38 AM) William Miller (MaCT USA): We need a way to discribe the resources and whether it is an INPUT or and OUPUT to the resource
(10:39 AM) Dr. Ravi Sharma: and ontologies do not necessarily follow hierarchies?
(10:39 AM) Dr. Ravi Sharma: except where there are parts of and in that case you can have them under taxonomy.
(10:40 AM) William Miller (MaCT USA): We need common identifiers for types of resource, data types, action required, etc.
(10:41 AM) Karen Guertler: OK, so since the v1 workplan, the workgroups have been somewhat reorganized...
(10:41 AM) Karen Guertler: and, it appeared that the Technology roadmap has interdependencies on the other three groups...
(10:42 AM) Karen Guertler: yet the drafts are due at the same time [finish-to-finish tasks]
(10:42 AM) Orit Levin (Microsoft): According to this (and I completly agree) that Use Cases are needed for both RA and Roadmap.
(10:42 AM) William Miller (MaCT USA): Service Broker will tie-in via Single Sign On (SSO) which is another NIST activity
(10:43 AM) Dr. Ravi Sharma: One more technical clarification about reflector, I have used NIST reflector to validate a software, format, standards, well formedness but lot

more complex functionalities as opposed to working group which deliberates, summary reflector for validation and working group for different deliberations, email, etc.
(10:44 AM) Orit Levin (Microsoft): How the documents should be submitted this week?
(10:45 AM) Dr. Ravi Sharma: upload authority?
(10:45 AM) Carl Buffington (USDA NRCS): Wo - would a charter per subgroup (draft) be a good deliverable for next wednesday's call?
(10:46 AM) Karen Guertler: I agree wrt cloud, and ALSO wrt other work being done wrt big data at, for example, NIH.
(10:48 AM) Carl Buffington (USDA NRCS): in addition ti NIH, at the end of this doc is more initiatives we should familiarize with and at least be aware of for our work. http://bigdatawg.nist.gov/WhiteHouse_big_data_press_release.pdf
(10:50 AM) Karen Guertler: That's a great observation re collaboration... though each team / subgroup might have a way of organizing their group's work.
(10:50 AM) Linda Pelekoudas: whatever method for collaboration is choosen should be documented in an operations record - over time people will come and go - having a reference will save time and help people to be up to speed faster
(10:50 AM) _James Kobielus_IBM: yes
(10:51 AM) Karen Guertler: @ Linda, I agree!
(10:51 AM) Linda Pelekoudas: expectations for what can be accomplished should be realistic - this is a short time window - and discussion in WG's takes time
(10:52 AM) Linda Pelekoudas: I am thinking that realistic on these objectives is to have a good definition of what the problem / needs are and some progress toward informationm gathering
(10:53 AM) Karen Guertler: Yes
(10:53 AM) Dave Raddatz (SGI): not yet
(10:53 AM) _James Kobielus_IBM: not yet
(10:53 AM) Karen Guertler: See the Submit New Input Doc screen.
(10:53 AM) Karen Guertler: appears that userid is required
(10:53 AM) ARC joined.
(10:53 AM) Linda Pelekoudas: I recall that a incomplete definition resulted in a 6 week delay in progress on the RA for Cloud - it wa epic
(10:53 AM) Virginia_Ross_AFRL/RCMT disconnected.
(10:54 AM) Linda Pelekoudas: The wiki used in the cloud work was unsupported which is why permissions to manage pages was challenged
(10:55 AM) Karen Guertler: collaboration via wiki, SharePoint site, or any other collaboration platform is preferable to collaborartion via e-mail. imo.
(10:55 AM) Linda Pelekoudas: Re Karen - I agree - it was a frustration
(10:56 AM) Linda Pelekoudas: train us -we will write it up
(10:56 AM) Dr. Ravi Sharma: No until you give priority or else tracking has to be ON for each type to know who revised what?
(10:56 AM) Scott Brim (Internet2): IMHO all groups should be able to _see_ all documents. Editing rights can be reserved to the subgroup.
(10:57 AM) Linda Pelekoudas: Re Scott - I second that notion
(10:57 AM) Linda Pelekoudas: it helps to promote cross group awareness
(10:58 AM) Karen Guertler: And, clear objectives.
(10:59 AM) Dr. Ravi Sharma: not possible to do without traceability?

(10:59 AM) Karen Guertler: Collaboration - SharePoint, wikis, similar
(10:59 AM) Dr. Ravi Sharma: hard to find such tools?
(10:59 AM) Linda Pelekoudas: Re Dr Sharma - Not so - there are many - the challenge is to choose and use in a common way
(11:00 AM) Linda Pelekoudas: Karen has listed several - I add that IBM had a whole division in software group to accomplish such interactions
(11:01 AM) Janis Beach (IMS): Consistent across groups is essential so that people on multiple groups don't have to learn mutiple tools.
(11:01 AM) Karen Guertler: I agree; and I also think that collaboration doesn't rely on any specific software... Just impt to have something other than e-mail.
(11:01 AM) Dr. Ravi Sharma: can we know name on one such tools where multiple people can without check-in and check-out?
(11:01 AM) Linda Pelekoudas: Domino
(11:01 AM) Karen Guertler: AND, folks can subscribe to receive posts / comments via e-mail.
(11:02 AM) Karen Guertler: :)
(11:02 AM) Linda Pelekoudas: let's start with what Wo has assembled - we can solve from there and add where needed
(11:02 AM) Pw Carey, (Compliance Partners, LLC): The email link connected us directly to this NIST conference....did your link work we're using Fire Fox
(11:03 AM) Dr. Ravi Sharma: domino is it freeware?
(11:03 AM) Linda Pelekoudas: it is very much a matter of desire to collaborate and cooperate
(11:03 AM) Carl Buffington (USDA NRCS): I have had good success with Confluence in managing multiple participants/authors on artifact development
(11:03 AM) Carl Buffington (USDA NRCS): http://www.atlassian.com/software/confluence/overview/team-collaboration-software
(11:03 AM) Linda Pelekoudas: No Domino is not free - but then that was not part of your requriement ;-)
(11:03 AM) Dr. Ravi Sharma: yes i agree but do want the product changes to be traceable?
(11:04 AM) Pw Carey, (Compliance Partners, LLC): Will there be a follow up e-Mail sent out to those attending today's meeting...?
(11:04 AM) Karen Guertler: Next steps?
(11:05 AM) Gururaj Pandurangi: will all email threads be shared on the current collaborative website?
(11:05 AM) Linda Pelekoudas: Could a website be donated?
(11:05 AM) Pw Carey, (Compliance Partners, LLC): SOAP HUB works for uploading documents...
(11:05 AM) Linda Pelekoudas: what if someone had one that could be offered?
(11:05 AM) Scott Brim (Internet2): Gururaj: imho mail should be archived on the site, but I prefer mail for text exchanges more than wiki forums.
(11:05 AM) Karen Guertler: Pretty easy to provision a SharePoint site. Whatever NIST prefers!
(11:07 AM) Dr. Ravi Sharma: can you send us the link?
(11:08 AM) Pw Carey, (Compliance Partners, LLC): When is the next meeting....?

(11:09 AM) Linda Pelekoudas: Yes weekly subgroup meetings are required to make progress

(11:10 AM) Karen Guertler: I'm still quite interested in how this initiative relates to other big data intiatives , especially wrt NIH.

(11:11 AM) Linda Pelekoudas: Re Karen - I agree - That is part of what each charter needs to add and take into consideration

(11:11 AM) Karen Guertler: as mentioned here: http://bigdatawg.nist.gov/WhiteHouse_big_data_press_release.pdf

(11:11 AM) Karen Guertler: @Linda, thanks.

(11:12 AM) Linda Pelekoudas: are all the web connects muted?

(11:14 AM) Linda Pelekoudas: a summary of state of the industry will be useful for comparison and contrast

(11:15 AM) Carl Buffington (USDA NRCS): Wo: is the vision that agencies could adopt (and possibly modify) the target architecture for their big data needs rather than start from scratch? When all deliverables are completed will they get presented to Government, Industry, etc?

(11:15 AM) vnavale(NARA): It appears that envisioned reference architecture is bottom up and not top down as is the case with the Federal EA framework?

(11:16 AM) Keith W. Hare (JCC Consulting, Inc.): The references in the White House Big Data press release are over a year old now. It would be useful to have more recent references to things like the NIH big data efforts.

(11:20 AM) Linda Pelekoudas: But the question remains - who are the projected users of the deliverables and what are thier needs

(11:20 AM) Carl Buffington (USDA NRCS): Wo: thanks makes sense.

(11:23 AM) Linda Pelekoudas: I think it was important that we understood what the use was when we defined the roles in Cloud - We need to understand that here

(11:23 AM) Karen Guertler: imo, if we don't know the potential end users, ... how can we establish the value of this effort?

(11:23 AM) Linda Pelekoudas: I can't get my audio to trun on in the web interface

(11:23 AM) _Nancy_Landreville_(_UnivMD_): I think WebEx is an easy product to use and is compatible with most govt systems whereas this venue is bandwidth intensive causing issues.

(11:24 AM) Scott Brim (Internet2): Linda: mute the web tool and dial in

(11:24 AM) Linda Pelekoudas: thanks scott I think that is the way - sigh

(11:24 AM) _Nancy_Landreville_(_UnivMD_): In creating various NIST standards, the Wiki was a good venue for updating information. The SAJAAC was a good excample with Alan Sill at the realm.

(11:25 AM) _Nancy_Landreville_(_UnivMD_): I participated in writing many ref architecture standards for NIST. This worked very well in the NIST venue.

(11:26 AM) William Miller (MaCT USA): we to align resource via an identfiication system which we currently have a new Global Unique Identifier for the Internet of Things (IoT)

(11:26 AM) William Miller (MaCT USA): The identifier can be used to rout the request/replies to a resource via the Service Broker

(11:26 AM) _Nancy_Landreville_(_UnivMD_): We created virtual labs for teaching students at the university. These cloud labs had a small learning curve but with a tutorial it worked well.

(11:26 AM) William Miller (MaCT USA): This is a decentralized distributed approach making the web one giant resource for big data.

(11:27 AM) William Miller (MaCT USA): it is importnat ot look at semantics and align to Web 3.0

(11:30 AM) Pw Carey, (Compliance Partners, LLC): But you have check when you initially sign in to have your words recorded.....

(11:30 AM) Karen Guertler: We hear ya.

(11:31 AM) Janis Beach (IMS): Need a unified conferencing tool like webex that does all together. Separate mechanisms are problematic in a group of this size.

(11:32 AM) Pw Carey, (Compliance Partners, LLC): Please include us in the next meeting, and our best wishes going forward....Respectfully yours, Pw Carey & thanks and have a good day way back East.....

(11:33 AM) Pw Carey, (Compliance Partners, LLC): Same here.....

(11:33 AM) Anil Srivastava (OHSL/ICTBioMed): Is participation in this working group restricted to US individuals and institutions?

(11:34 AM) Linda Pelekoudas: Not normally Anil - in other groups there were many spread out in different countries and interests -bear in mind that rules of open and public apply

(11:38 AM) _Nancy_Landreville_(_UnivMD_): We needed to be up and running. Understand..

(11:38 AM) Karen Guertler: Just fyi, I've attended other mtgs re Adobe Connect in Federal. This venue was fine.

(11:38 AM) _Nancy_Landreville_(_UnivMD_): Yes Adobe Connect is good.

(11:39 AM) Janis Beach (IMS): Makes sense. Thank you for the clarification. If it can come online later, it might be beneficial to the collaboration.

(11:44 AM) Keith W. Hare (JCC Consulting, Inc.): We don't need a complete definition of "Big Data" -- we only need a good enough definition to understand the other pieces.

(11:45 AM) _Nancy_Landreville_(_UnivMD_): And...a means of consensus.

(11:46 AM) Scott Brim (Internet2): right, definition based in use cases

(11:46 AM) Karen Guertler: Good points from Keith, Nancy, Scott above...

(11:47 AM) Karen Guertler: and, lots of great dialog from the NIST January workshop on Cloud and big data.

(11:52 AM) Bob Marcus (ET-Strategies): In September deliverables, we can list consensus informations and areas where there are competing alternatives. Discussions after September can help resolve among alternatives.

(11:52 AM) Susan Malaika(IBM): will the (cleaned up) chat be saved as part of the meeting minutes?

(11:52 AM) Orit Levin (Microsoft): also on WEB

(11:53 AM) Linda Pelekoudas: Re Bob - yes that is a good path forward

(11:54 AM) Karen Guertler: I agree the chart on screen is helpful. However, does this reflect the *current* organization of subteams and 'reflectors'?

(11:54 AM) Bob Marcus (ET-Strategies): To Susan: I am capturing the chat and will add them after cleansing to the minutes.

(11:55 AM) Karen Guertler: That is, while I completely agree with the emphasis on use cases, roles, etc.... I got the sense that the overall initiative has been reorganized since this v1 timeline was scheduled.

(11:56 AM) Karen Guertler: And, the tech roadmap appears dependent upon the three other workstreams - definitions, taxo, and ref arch.
(11:56 AM) PavithraKenjige PK Technologies: There is interdependancy between the out put of each one of these items on the columns
(11:56 AM) Karen Guertler: with all drafts due the same week.
(11:56 AM) PavithraKenjige PK Technologies: I agree with her
(11:57 AM) Karen Guertler: @ PavithraKenjig I agree.
(11:57 AM) PavithraKenjige PK Technologies: For planning and for deliverable dates on has to consider the interdependency between these actiivties
(11:57 AM) Orit Levin (Microsoft): Agreed :-)
(11:58 AM) Linda Pelekoudas: I expect that where there is dependancy on a up stream work group that a clear identification of what is needed would give enough structure of information to know what is needed and from where
(11:59 AM) Scott Brim (Internet2): I believe IPv4 vs v6 is a level lower than this group should care about.
(11:59 AM) Linda Pelekoudas: good practice to keep to meeting times
(11:59 AM) PavithraKenjige PK Technologies: A sequence plan between activities is needed
(12:00 PM) Linda Pelekoudas: Well this was fun - talk to you all again next week if not before
(12:00 PM) William Miller (MaCT USA): Big Data is agnostic to IPv4 or IPv6
(12:00 PM) _Nancy_Landreville_(_UnivMD_): Updated and sent you an email. Will assist co-chair for Security and Privacy.
(12:01 PM) Karen Guertler: Thanks!
(12:02 PM) Karen Guertler: Would you please identify the types of use cases needed? that would be very helfpul.
(12:02 PM) Karen Guertler: Use cases have different meaning in different contexts. Thks!
(12:31 PM) Bob Marcus (ET-Strategies): Primary customers should be procurement in government and industry. Working Deliverables are an FYI. Secondary customers are solution providers from industry, open source, and academia. Working Group deliverables provide gap analysis to help guide future activities.